DOCUMENT RESUME

ED 129 915 TM 005 789

AUTHOR Spool, Mark D.
TITLE Performing a Content Validation Study.
PUB DATE Mar 75
NOTE 16p.; Paper presented at the Annual Meeting of the Southeastern Psychological Association (21st, Atlanta, Georgia, March 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Content Analysis; Correlation; Evaluation; Guidelines; Item Analysis; *Job Analysis; Test Construction; *Test Validity
IDENTIFIERS *Content Validity

ABSTRACT

Content validity is concerned with three components: (1) the job content; (2) the test content, and (3) the strength of the relationship between the two. A content validation study, to be considered adequate and defensible should include at least the following four procedures: (1) A thorough and accurate job analysis (to define the job content); (2) A thorough, intensive review and evaluation of the test by field experts to determine its content and relatedness to a job characteristic at the appropriate level. The use of field experts is important since content validity relies heavily upon expert judgment; (3) A comparison between the test content and the job content considering many important factors, to assess the extent to which the total test is job related; and (4) Documentation of each procedure of the content validation study. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED129915

TM005 789

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERFORMING A CONTENT VALIDATION STUDY

Mark D. Spool
S.C. Personnel Division, Research Unit

For my presentation, I shall describe the essential components of a content validation study and how a good study might be performed. Discussion will be limited, however, to studies involving written and performance tests because of their greater applicability to content validation. Emphasis will be placed upon reality, practicality and other areas not frequently emphasized in the literature and in presentations such as this one.

Before proceeding to a discussion of performing a content validation study, I should like to comment generally about content validation. It is often proposed that content validation not be considered as an end in itself but rather as an intermediary step which must be substantiated later by criterion-related validity. Practically speaking though, content validation of a test should be performed as if it were an end in itself such that it can stand alone. The chances are, you see, in most settings, a follow-up criterion-related validity study is not likely to be performed because it may not be considered technically or administratively feasible.

Content validation has often been viewed as an easy validation process which involves relatively few requirements. In fact, many people seem to confuse content validity with face validity. To the contrary, the content validation process involves somewhat more than a perusal of the test content and a superficial determination of the relationship between the test content and the job content. It involves many carefully planned interrelated operations or procedures. The soundness and defense of each procedure affects the soundness and defense of the whole content validation process.

Paper presented at the Southeastern Psychological Association symposium on content validity, Atlanta, March, 1975.

Presently at Michigan State University

Content validation operations might be utilized in two situations: (1) when developing a test and (2) when reviewing an existing test. The former utilizes content validity as a "built-in" process where test items are developed or carefully selected during test construction on the basis of their job relatedness. The resulting test, therefore, is content valid. The latter, on the other hand, examines a pre-developed test to assess the degree of content validity.

The advantages of considering content validity as a "built-in" process, rather than an "after-the-fact" process, are clear. Developing tests with built-in validity usually does not require significantly more time nor effort than content validating pre-existing tests since both require a job analysis, a review and evaluation of the test content and an assessment of the relationship between the test and the job. The built-in-process yields a test that is valid and perhaps of better quality. If a test, however, were developed without regard for the job relatedness of each item and then assessed for content validity, major problems are liable to be encountered. If it were determined, for instance, that the test is weakly related and/or poorly proportioned to the job content, then any claim of validity would be weak. The weaknesses of the test, then, have been revealed and the test specialist may well be obligated to correct them as early as possible -- a circumstance which could present an unwarranted burden. The additional time now required could well have been avoided if the built-in process were used. Moreover, the additional effort would be approximately equal to that involved in developing a test with built-in content validity.

COMPONENTS OF A CONTENT VALIDATION STUDY

In order to identify the components of a content validation study, the definition of content validity must be examined. Content validity is the assessment of the degree to which the content of the test is directly related to the content of the job. As stated in the 1974 Standards for Educational and Psychological

Tests and Manuals (hereinafter "APA Standards"), prepared by a joint committee of the American Psychological Association, the American Education Research Association, and the National Council on Measurement in Education, "To demonstrate the content validity of a set of test scores, one must show that the behaviors demonstrated in testing constitute a representative sample of behaviors to be exhibited in a desired performance domain."[1]

Accordingly, content validity is concerned with three components: (1) the job content (behaviors to be exhibited in the performance domain), (2) the test content (behaviors demonstrated in testing) and (3) the strength of the relationship between the two. These three components involve several interrelated procedures. A content validation study, in order to be considered adequate and defensible, it is believed, should include at least the following four procedures:

1. First, a thorough and accurate job analysis (to define the job content).
2. A thorough, intensive review and evaluation of the test by field experts. This procedure should include an analysis of each test item to determine its content and relatedness to a job characteristic at the appropriate level. The use of field experts is important since content validity relies heavily upon expert judgment.
3. A comparison between the test content and the job content, considering many important factors, to assess the extent to which the total test is job related; that is, content valid.
4. Documentation of each procedure of the content validation study (to substantiate claims of content validity).

These procedures, however, are based upon several important assumptions which must be appropriately met and adequately substantiated. For example, it is assumed that the field experts really know the job content and job requirements; that the job analysis results represent a true picture of the job; and that the test items accurately measure the job content area to which they are related. Awareness of

these assumptions and the thoroughness and accuracy of each procedure will help to assure that the assumptions are met and well substantiated. Each procedure, then, will be discussed in detail.

DESCRIPTION OF A CONTENT VALIDATION STUDY

Job Analysis. The first step in performing any content validation study, whether developing a content valid test or content validating an existing test, is the job analysis. In order to claim that a test is job related, one must first study the job and know its content. Furthermore, EEOC requires that evidence for content validity be accompanied by sufficient information from job analysis to demonstrate the relevance of the test content.

Job analysis describes a method of studying a job in order to identify and describe its content. There are many job analysis methods, each providing different kinds of information about the job. For content validation purposes, it is necessary to use a job analysis method which yields the essential knowledges, skills, abilities (KSA's) or behaviors (tasks) comprising the job in question.

Several job analysis methods exist which identify those KSA's or tasks which are required to perform the job. Among them are:

1. Functional Job Analysis (FJA), which describes a job by its tasks in relation to data, people, and things.
2. Critical Incident Technique (CIT), which describes a job in terms of concrete and specific behaviors, characteristic of effective and ineffective workers.
3. Job Element Approach, which describes a job in terms of its component elements (KSA's or personal characteristics) that are required for successful job performance.
4. Domain Sampling, which describes a job by identifying and relating tasks and KSA's required to perform the job.

Use of field experts. Job analysis has been defined as the "coming together of subject-matter-experts who provide content input and psychometric experts who

construct an examination using that input."[2] In order to thoroughly and accurately analyze a job, the data must come from a group of people well familiar with the job and its requirements. These people, called subject-matter-experts or field experts, are usually superior incumbents, although an immediate supervisor recently promoted from a position in that job may suffice. The psychometric expert or test specialist who is responsible for the construction of the test must also be fully involved in the job analysis.

Field experts are essential to the content validation process since claims of content validity rely heavily upon expert judgment. It is important, therefore, to utilize them throughout the whole study. In every case, the expertise of each field expert demonstrated by education, training, and relevant work experience should be established and thoroughly documented.

Determination of appropriate type of test and job analysis method. As previously stated, regardless of whether one is developing a content valid test or assessing the content validity of an existing test, a job analysis must first be performed. In other words, in order to know what to look for in an applicant (via a test), the employer must first know what the job consists of and its requirements (via a job analysis).

There are several types of tests which lend themselves to content validation. The majority of tests fall into one of three types: (1) work sample (in which an actual part of the job becomes part of the testing situation), (2) simulation (approximations of work samples from the job), and (3) "paper and pencil" or written test (which measures the knowledges, skills and abilities required to perform the tasks found in the job). A performance test may be a work sample or a simulation, depending upon the content of the test and the conditions of the test administration.

The most desirable type of test is the work sample because it is a direct measure of the job and, therefore, requires relatively little inference in relating test content to job content. A simulation is a less direct measure of the job than a work sample and a written test is the least direct measure of the three. The more

direct the measures of the job, the stronger the statements of content validity.

Often neither a work sample nor a simulation is practical to use because both require considerable time and expense to administer to an applicant. Because of this, employers tend to use written tests which can be administered to many applicants simultaneously with much less difficulty and expense. The use of a written test is acceptable, of course, provided there is sufficient documentation from job analysis to support the job relatedness of the test.

As has been indicated, there are several acceptable methods of job analysis and several types of tests. The methods of job analysis vary in the kinds of data they yield. For example, some job analysis methods yield tasks, others yield KSA's and at least one job analysis method yields both kinds of data. In much the same manner, the types of tests may vary in the different job characteristics which they attempt to measure. Whereas work samples and simulations are concerned with tasks, written tests measure the KSA's required to perform tasks. Because of the variations in data, it is essential that the test specialist assure that the data from the job analysis is compatible with the type of test being validated.

The interrelationship between the job analysis method and the type of test being validated is illustrated by the different situations in which content validity is examined. When assessing the content validity of an existing test, which was probably neither developed nor selected on the basis of a job analysis, the type of test (i.e., whether it measures KSA's or tasks) suggests the method of job analysis which should be performed. If, for example, a written test were involved, then the job analysis method chosen for the content validation process should yield KSA's required in the job.

On the other hand, in developing a content valid test, the results of the job analysis (i.e., KSA's or tasks) help to identify the type of test which should be used. If the job analysis produces tasks, for example, then a work sample or simulation should be considered. Caution is in order, however, if the original job analysis produced only tasks and, for practical reasons, only a written test can be

used. In this case the test specialist must produce and document a defensible set of KSA's which are required to perform the tasks. The KSA's, carefully related to the tasks, will then serve as the basis for developing the written test and for demonstrating job relevance. For this reason, it is best to begin with a job analysis method (or combination of methods) which will yield both KSA's and tasks. Furthermore, stronger claims of validity are warranted since less inference is required when test items, measuring KSA's, are linked to the job through tasks performed in the job.

Job Content. A significant issue regarding job content is which KSA's or tasks should be or can be tested. Of particular concern here is practicality. Not all characteristics of the job need to be or can be measured by a test. The APA Standards and the EEOC Guidelines on Employee Selection Procedures, however, require that a test include all, or nearly all, important parts of the job.

Frequency and criticality data can be used to determine which KSA's or tasks should be tested. Frequency refers to the number of times a task is performed in the job or a KSA is required in order to perform the job. Criticality is concerned with the consequences of inadequate performance of a task or the consequences of the absence of a KSA. Tasks or KSA's which are frequently required and/or are critical should be considered for testing. Tasks or KSA's which are neither frequently required nor critical, on the other hand, do not necessarily need to be chosen as a part of the job content to be tested. Precaution should be taken, however, to include very critical KSA's or tasks which may be only infrequently required or performed.

Two other factors should be considered in determining which KSA's or tasks to test for: (1) the feasibility of measurement and (2) utility. Personality characteristics, such as honesty, for example, may not be feasible to test for because they often cannot be validly and reliably measured. Utility becomes a concern when almost everyone or almost no one has a required KSA or can perform a task. In such cases, it is usually not worthwhile to test for the KSA or task unless it is

a critical factor which must be present for minimally acceptable performance of the job. Color-blindness, for example, is possessed by very few people; but it may be important to consider in hiring applicants for certain jobs, such as artist, which requires an employee to distinguish between colors.

Knowing which job characteristics to test for, then, is very important. Of equal importance, however, is knowing how much a certain characteristic should contribute to the total test score; that is, knowing the relative proportions of the job characteristics which are to be tested. This information is essential to statements of validity. To paraphrase the EEOC Guidelines, to demonstrate content validity, one must show that the test content is a representative sample of the job content (as identified by the job analysis). If a secretarial job, for example requires 90% typing and 10% filing, then a test for that job must measure those job characteristics in those proportions. If a test for that job, however, measured 10% typing and 90% filing, then it would not be considered valid.

As a matter of interest, the job characteristics (KSA's or tasks) finally chosen for testing may influence the effectiveness of a test. For instance, if critical or frequent KSA's or tasks are omitted or poorly proportioned, then applicants most likely to be successful on the job may not score high and, thus, not be selected. To help assure that the test will be effective, prerequisite KSA's or tasks as well as the more important characteristics of the job which distinguish between strong applicants and weak applicants should be identified and included in the test.

The appropriate emphasis of each of the job characteristics can be determined only by job analysis and must be independently established for each test. Unfortunately, most job analysis methods do not provide this information and there exist no definite guidelines specifying a procedure to use to determine proportionality. Frequency and criticality data are often used. Whatever approach is taken, however, it should yield relative weights or percentages for each KSA or task to be measured by the test. If the job analysis method chosen does not yield this information, the technique, then, must be modified to produce proportionality or weighting data.

Deriving KSA's and tasks from job analyses and deciding upon which ones to test for and in what proportion involves judgment. Because of this, it is very important to document the procedures followed in the job analysis: description of the method used, qualifications of the people involved, instructions given and the results of the job analysis.

Review and Evaluation of Test Content. Once the job analysis has been performed, the content of the test can be analyzed. The original field experts may be retained or different people may be used in this procedure. If new field experts are used, however, it would be best for them to become familiar with the results of the job analysis before analyzing the test content.

The analysis of the test content should be performed with great precision, carefully reviewing and evaluating the test content in relation to the job content. The purpose, of course, is to assess the job relatedness of the test. Because assessment of the job relatedness of work samples and simulations is fairly straightforward, the focus of this section will be on written tests. With respect to written tests, then, each item of the test must be analyzed individually.

The review and evaluation of the test content is very important. A brief or cursory inspection of the test will not suffice. While a test item may on the surface appear to be related to or measuring one thing, it may in fact be related to or measuring something entirely different. In other words, content validation involves more than a casual review of test items. It requires a careful scrutinization and serious evaluation of each item.

As part of this scrutinization, the content area of each test item must be determined. Content area refers to the subject matter or category of the KSA or task which the test item is attempting to measure. Some examples are: unstructured interviewing techniques; four basic arithmetic operations; and legal terminology. Identification of the content area of an item should be made and agreed upon by the field experts.

Relationship between test content and job content. Once the content area of a test item has been determined, its relationship to the job can then be established. To do this, the content area of the test item must be compared to and matched with one or more of the KSA's or tasks identified in the job analysis. Typically, in content validation studies, test items are considered in relation to either KSA's or tasks, but not to both. Moreover, the items of written tests, because they measure KSA's are almost always compared only to KSA's. This may be partly due to conceptual differences between KSA's and tasks. More likely, however, it is probably because most job analysis methods produce either KSA's or tasks, but not both.

The defensibility of any test, though, rests with its job relatedness. Tasks are easily recognizable characteristics and they are closely and directly related to the job. For these reasons it would seem best to link test items to tasks. But, written tests measure KSA's. It would be wise, then, for the sake of defensibility and soundness of the study, to evaluate test items in relation to tasks through the KSA's required to perform those tasks. The most likely time to establish the necessary linkage between tasks and KSA's would be during the job analysis. By using such a procedure the test is being linked more directly to the job and, therefore, stronger claims of validity may be warranted. Moreover, the information made available by this approach should greatly facilitate the difficult task of item writing.

Ratings by the field experts can be utilized to estimate the extent of relationship between the test and the job. Each test item should be rated according to its degree of relevancy to the KSA or task identified as being related. Such ratings are important for they provide the test specialist with an indication of the validity of a test -- the more job related the test items, the more content valid the test. If ratings by field experts are used, then some measure of interrater reliability should be determined and documented.

There are two important factors which should be considered in the rating of the job relatedness of test items: (1) the extent of agreement between the content area of the test item and the content area of the job characteristic (KSA or task) and (2)

the degree to which the level of the job characteristic measured by the test item is equivalent to the level of that job characteric found in the job. Almost all content validation studies take into consideration the former factor. It is straightforward and, therefore, easy to deal with when rating test items. Consideration, too, should be given to the total job, including the kind of work, kind of people dealt with and setting. The latter factor has often been overlooked or not sufficiently emphasized. In order for a test item to be related to a job characterisitc, that is, be content valid, not only must the subject category (content area) of both the item and the job characteristic be the same but the levels, too, must agree. Level refers to a certain degree of difficulty or a particular rank of ability or achievement.

The following may serve as an example of the importance in considering the agreement between the levels of the job characteristics tested and those found in the job. If a job involves calculating payments due and requires a knowledge of the four basic arithmetic operations, then a test item may contain addition, subtraction, multiplication, or division. If the test item, however, measured arithmetic at a higher level, such as an intricate combination of the four arithmetic operations, and, if the job did not require such an arithmetic operation, then that item would be less job related -- perhaps not job related (content valid) at all. Such an item may also be answered incorrectly by many qualified applicants, thus causing the test to be ineffective in differentiating more qualified from lesser or unqualified applicants.

It is important to point out here one of the major shortcomings of content validation -- that is, it does not guarantee the effectiveness of a test. Content validation may help to assure effectiveness by requiring test items to be proportionately related, considering both content and level, to the job characteristics. Consequently, the more qualified applicants should answer more test items correctly than the lesser qualified applicants. A test, however, may be content valid regardless of whether it differentiates the more qualified from the lesser qualified applicants. The main reason for this is that the effectiveness of a test is affected largely by the way the questions or the choices of response are worded. A test, even though job related, may

not be effective if the questions are too vague, difficult or easy, or if the choices of response can be readily deduced. In such cases, almost everyone may score very low or very high on the test. Consequently, the applicants are not being differentiated.

In summary, in order to assess the content validity of a test, each individual item must be carefully scrutinized and evaluated with regard to at least the two factors discussed. Accomplishment of such an intensive review and evaluation of the test items may require the field experts to take the test themselves, just as an applicant would. In this way, the field expert can more effectively study each individual test item, including the question, choices of response and the appropriateness of the correct response.

The test-item review session must be thoroughly documented. Documentation should include the qualifications of the field experts (if not already obtained), the instruction given to them, their evaluation of the job relatedness of each test item, the reliability of any ratings and complete descriptions and treatments of data produced (such as averaging).

Overall relationship between test and job. "Evidence of content validity is required when the test user wishes to estimate how an individual performs in the universe of situations the test is intended to represent."[3] Accordingly, the content validity of a test can be interpreted as the degree to which the content of the test is related to the content of the job which the test claims to measure. In situations where more than one test is used for selection in one job, the content validity of each test (or sub-test) may vary, depending upon the extent to which the items are related to the job content area(s) which each test is intended to measure. Whether one or more than one test is used, however, the important KSA's or tasks required in the job must be measured. It must not be forgotten that the test content not only needs to be related to the job content, it must be related proportionately.

ASSESSMENT OF CONTENT VALIDITY

Assessment of the content validity of a test is not simple. This is probably

because there are no set guidelines stating all that is required to demonstrate content validity. Content validity is also dependent upon judgment. Like construct validity, there is no quantitative index of content validity. Content validity is evaluated by the accumulation of many evidences surrounding the test. The strength of a claim of content validity depends upon the strength of each of these evidences in demonstrating content validity. Almost all of the information obtained in the content validation study may serve as evidence to support the content validity of a test. The use of these evidences in the assessment procedure demonstrates their importance to the content validation process.

The information utilized from the content validation study is:

1. The degree of relevancy of the test items to the job characteristics. This takes into consideration:
   a. the degree of agreement between the content of the KSA's or tasks measured by the test items (test content) and the content of KSA's or tasks performed in the job (job content) and
   b. the degree of agreement between the level of KSA's or tasks tested and the level of KSA's or tasks required in the job.
2. The extent to which the content of the test is proportionately related to the content of the job.
3. The extent to which all the essential KSA's or tasks required in the job are samples by the test or tests.

SUMMARY

In summary, a content validation study consists of many interrelated procedures, each one important to the defense of the total content validation process. The key to content validation is to pay particular attention to each procedure, assuring its soundness and documenting the essential data. The extent to which each of these evidences demonstrates the job relatedness of a test determines the strength of any claim of content validity.

ONE LAST COMMENT

Some recent court cases have had direct impact upon content validation. One case in particular, Edward L. Kirkland, et al., Plaintiffs v. New York State Department of Correctional Services, et al., Defendants (commonly called the Kirkland case) has several implications affecting the procedures followed in performing a content validation study. Apparently a job analysis should be performed prior to the development of each examination and the critical or more important characteristics of the job which differentiate between more qualified applicants from those less qualified must be included in the examination. These job characteristics further must be weighted in accord with their importance to the job. Field experts, to be considered qualified, apparently must be of appropriate rank and sufficiently experienced in the current skills, knowledges, and abilities required in the job for which the test is developed.[4] The procedures outlined in the present paper, it is believed, may adequately meet the concerns of the courts and, at the same time, provide the test specialist with a basic plan for content validating tests.

NOTES

1. American Psychological Association, Standards for Educational and Psychological Tests, Washington, D.C., APA, 1974, p. 28.

2. Lasker, Edward L. Kirkland, et al., Plaintiffs v. New York State Department of Correctional Service, et al., Defendants, U.S. District Court, Southern District, New York, 1974.

3. American Psychological Association, Standards for Educational and Psychological Tests, Washington, D.C., APA, 1974, p. 28.

4. Wisner, R.W., "The Kirkland Case - Its Implications for Personnel Selection," Public Personnel Management Journal, vol. 4, (4), 1975.